IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received February 19, 2020, accepted April 1, 2020, date of publication April 3, 2020, date of current version April 20, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2985453

A Face Spoofing Detection Method Based on Domain Adaptation and Lossless Size Adaptation

WENYUN SUN[1,2,3], (Member, IEEE), YU SONG[1,2,3], (Member, IEEE), HAITAO ZHAO[4], AND ZHONG JIN[5]

[1]College of Electronics and Information Engineering, Shenzhen University, Shenzhen 518060, China
[2]Shenzhen Key Laboratory of Media Security, Shenzhen University, Shenzhen 518060, China
[3]Guangdong Key Laboratory of Intelligent Information Processing, Shenzhen University, Shenzhen 518060, China
[4]School of Information Science and Engineering, East China University of Science and Technology, Shanghai 200237, China
[5]School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China

Corresponding author: Wenyun Sun (wenyunsun@szu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61902250, and in part by the China Postdoctoral Science Foundation under Grant 2018M643183.

ABSTRACT In this paper, a face spoofing detection method called the Fully Convolutional Network with Domain Adaptation and Lossless Size Adaptation (FCN-DA-LSA) is proposed. As its name suggests, the FCN-DA-LSA includes a lossless size adaptation preprocessor followed by an FCN based pixel-level classifier embedded with a domain adaptation layer. The FCN local classifier makes full use of the basic properties of face spoof distortion namely ubiquitous and repetitive. The domain adaptation (DA) layer improves generalization across different domains. The lossless size adaptation (LSA) preserves the high-frequent spoof clues caused by the face recapture process. The ablation study shows that both DA and the LSA are necessary for high-accuracy face spoofing detection. The FCN-LSA obtains competitive performance among the state-of-the-art methods. With the help of small-sample external data in the target domain (2/50, 2/50, and 1/20 subjects for CASIA-FASD, Replay-Attack, and OULU-NPU respectively), the FCN-DA-LSA further improves the performance and outperforms the existing methods.

INDEX TERMS Domain adaptation, face anti-spoofing, face liveness detection, face presentation attack detection, face spoofing detection, forensics, machine learning, pattern recognition.

I. INTRODUCTION

Faces can be captured conveniently by digital cameras, web cameras, smart phones, etc. The convenience is a double-edged sword. It makes faces become not only the most widely used but also the most untrustful biometric modality. With the fast development of face recognition, the modern face recognition algorithms [1]–[3], especially deep networks trained on large scale datasets, can surpass human performance, but they may be easily fooled by face spoofing attacks which can be easily launched by inexperienced attackers. The security of the face modality become an important and practical problem in these days. Attackers use photos, video records, 3D plastic masks, etc. to mimic genuine faces, fool the face recognition algorithm and get the unauthorized system access. On the defending side, the face spoofing detection, a.k.a. face anti-spoofing, face liveness detection, or face presentation attack detection, is an auxiliary task for securing the face verification systems. It classifies the detected faces into two types: the genuine and the spoof. The spoof faces with spoof clues including Moiré patterns, certain Local Binary Patterns (LBPs), and compression/color degradation, etc. will be rejected by the face spoofing detector before the face recognition. Recently, deep learning-based face spoofing detection has become popular. For example, some Convolutional Neural Networks (CNNs) [4]–[6] and some Fully Convolutional Networks (FCNs) [7]–[10] are used for process static frames. Some Recurrent Neural networks (RNNs) [7] are used for aggregating all the frames in the video. These methods learn the image features as well as the classifiers at the same time in an end-to-end scheme. This work is focused on improving the performance of the FCN-based face spoofing detection methods. The main contributions of this work include

- A new face spoofing detection method called the Fully Convolutional Network with Domain Adaptation and Lossless Size Adaptation (FCN-DA-LSA) is proposed. Its improvements are twofold. First, the domain adaptation layer is designed and embedded in the FCN

The associate editor coordinating the review of this manuscript and approving it for publication was Javier Medina.

to improve generalization across different domains. Second, the lossless size adaptation preprocessor preserves the high-frequent spoof clues caused by the face recapture process.

- The proposed method is empirically verified in the ablation study and compared with the existing methods. The ablation study compares four related methods including FCN, FCN-LSA, FCN-DA, and FCN-DA-LSA under the cross-dataset protocols on CASIA-FASD and Replay-Attack dataset. It shows that both domain adaptation and the lossless size adaptation are necessary for high-accuracy face spoofing detection. The proposed methods are further compared with 21 existing methods including DeepPixBiS [9], Auxiliary Supervision [7], Noise Model [11], STASN [12], Colour Texture [13], [14], IQM+IQA+SVM [15]–[17], KSA+DA+SVM [18], MMD-AAE [19], and MADDoG [20] under both the cross-dataset protocols and the standard protocols of OULU-NPU. The FCN-LSA obtains competitive performance among the state-of-the-art methods. With the help of small-sample external training data in the target domain, the FCN-DA-LSA further improves the performance and outperforms the existing methods.

The rest of the paper is organized as follows. Section II reviews the related work. In Section III, the main method is proposed. The experiments are conducted, and results are analyzed in Section IV. Finally, Section V gives the conclusions.

II. RELATED WORK

A. FACE SPOOFING ATTACKS

Attackers use photos, video records, 3D plastic masks, etc. to mimic specific genuine faces and fool the face recognition algorithms. These behaviors are all defined as the face spoofing attacks. The face spoofing attacks can be generally divided into three categories, namely photo attacks, video attacks, and 3D mask attacks [21], [22]:

- In photo attacks, attackers take a photo of an authentic person, or download a photo from social networking sites, then display the photo on various screens, or print it on a piece of paper. Without face spoofing detectors, face recognition methods usually cannot differentiate the genuine faces and the spoof faces on screens/papers. Photo attacks are very easy to commit.
- To detect photo attacks, some face spoofing detectors use the dynamic clues such as eye blinking. Video attacks can pass such detectors. In video attacks, attackers record a video of an authentic person, or download a video from social networking sites, and then replay the video on various screens. Video attacks are more real than photo attacks. They are also not difficult to commit.
- 3D mask attacks are the advanced version of 2D photo/video attacks. In 3D mask attacks, attackers make a plastics mask or a silicon mask of an authentic person.

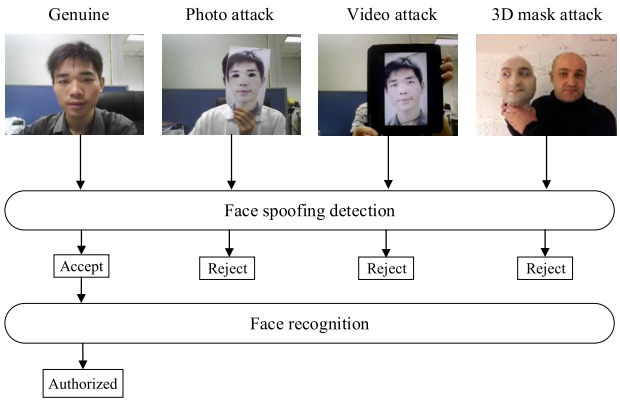


FIGURE 1. The face spoofing attacks and face spoofing detection task.

This kind of attacks is less prevalent as compared to the photo attacks and the video attacks since it is relatively difficult to make a mask. 3D mask attacks are usually not included in common face spoofing detection datasets.

Fig. 1 demonstrated the detection of the above mentioned as well as the recognition of a genuine face. The face spoofing detection is a kind of firewall to keep the face recognition system away from these attacks.

B. GENERAL FACE SPOOFING DETECTION

As shown in Fig. 2, the face spoofing detection methods are generally divided into two groups by their input types, namely static images or dynamic videos). The methods for images can be further divided into four categories, namely the frequency domain feature-based methods, spatial domain feature-based methods, deep feature-based methods, and transfer learning feature-based methods. For photo/video attacks, faces are first captured by camera, then printed/displayed on papers/screens are recaptured by another camera. This process is a special cases of the image recapture which has a wide variety of defending methods. The mathematic model of the images recapturing from an LCD screen was investigated by Muammar *et al.* [23]. Since the RGB sub-pixels are used to compose the color pixels on LCD screens, and the discrete pixels are recaptured by the discrete Bayer color filter mosaic in the camera, the recapturing process generates a special kind of textural artifacts, namely Moiré patterns. Similar to the LCD screen, the Moiré patterns can also be found on printed materials in which the halftone technique is used. Just as the name suggests, Moiré patterns is a periodic noise in the image. In the frequency domain, the Discrete Fourier Transform (DFT) is a useful tool for analyzing Moiré patterns [24], [25]. In the spatial domain, image features including LBP and Scale-Invariant Feature Transform (SIFT) can also be used [26], [27]. Another useful clue in the recaptured faces is the degradation of color and image quality, which can be represented by the color texture histogram features [14], the color moments feature [25], and image quality assessment features [16], [17]. Recently, the CNNs [4]–[6] and the FCNs [7]–[10] were used for directly

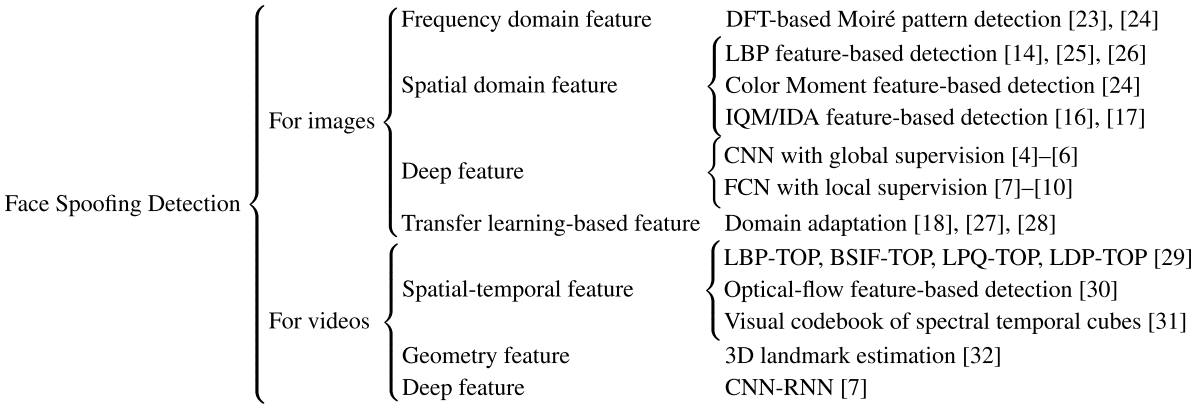


FIGURE 2. The overview of the existing face spoofing detection methods according to the target media and the extracted features.

learning the classifier as well as the features for face spoofing detection task. These deep learning-based methods will be reviewed in Section II-C–II-D in detail. Finally, some studies [18], [28], [29] were focused on the adaptation between different domains in which the input devices (cameras) and the output devices (screens/printers) are different.

As shown in Fig. 2, the methods for videos can be divided into three categories namely spatial-temporal feature-based methods, geometry feature-based methods, and deep feature-based methods. A straightforward way of detecting the spoofing artifacts in videos is devised by extending the 2D spatial domain features into 3D spatial-temporal features. De Freitas Pereira *et al.* [30] used LBP on Three Orthogonal Planes (LBP-TOP) features to detect face spoofing attacks. In a similar fashion, most histogram-based 2D features can be extended to their 3D counterparts (e.g., HOG-TOP, BSIF-TOP, LPQ-TOP, LDP-TOP). The spatial-temporal features are extracted in the same manner of the spatial feature. Besides, optical flow is a useful tool to extract motion in a video [31]–[33]. Yin *et al.* [34] investigated the optical flow to find the motion clues of face spoofing. Pinto *et al.* [35] proposed a face spoofing detection method based on a low-level motion feature and a mid-level visual codebook feature. De Marsico *et al.* [36] detected the facial landmarks and exploited geometric invariants for detecting replay attacks. Liu *et al.* [7] used the RNN to aggregate the features and depth maps predicted from single frames.

C. CNN-BASED FACE SPOOFING DETECTION

CNNs have become popular in many computer vision areas. Using CNNs to classify the cropped faces into genuine and spoof classes is a straightforward way. Menotti *et al.* [4] use meta-parameter search methods to find suitable CNN architectures for iris, face, and fingerprint spoofing detection, respectively. To limit the number of meta-parameters to be searched, their CNN only has three convolutional layers. Rehman *et al.* [5] trained an 11-layer VGG network with its two derivations for face anti-spoofing in an end-to-end scheme. Nagpal *et al.* [6] explored deeper ResNet and GoogLeNet for training the face spoof detector. In the above work, decision of the CNNs are based on the whole face crops. These CNNs are referred to as global classifiers and the whole face crops is called global supervision.

Since the spoof clues, including Moiré patterns, certain LBP patterns, and compression/color degradation, are ubiquitous and repetitive (see Section II-D). Feeding the whole face into CNNs is inefficient. Instead, we can feed small face patches to the classifier. For example, Atoum *et al.* [8] designed a patch-based CNN to detect the spoof patterns in extracted face patches of 96×96 pixels. This is the basic version of the local supervision which is the key to the high performance. But we think the basic one is inefficient since the pixels outside the small patches are wasted. In the next section, the local supervised FCN-based face spoofing detection methods will be reviewed. They can avoid the problem of data inefficiency.

D. FCN-BASED FACE SPOOFING DETECTION

The spoof distortion is a kind of high-frequency weak signal added to the clean face image. Based on the case study of Jourabloo *et al.* [11], there are two basic properties of spoof distortion namely **ubiquitous** and **repetitive**. First, the ubiquitous property makes the distortion exists everywhere in the spatial domain. Second, the repetitive property makes the distortion be a spatial repetition of certain regular patterns. Thus, it is sensible to use FCNs to model the mapping from the local patches to the local labels. The local labels located in the same face form a map of all ones or zeros. For example, George *et al.* [9] and Sun *et al.* [10] give a general theoretical analysis to demonstrate that the local labels are more suitable than the global labels for face spoofing detection. Pixel-level local ternary labels are employed to train the FCN which achieves state-of-the-art performance. Besides, Liu *et al.* [7] and Atoum *et al.* [8] use depth map as the auxiliary labels to train their FCN. The depth map is similar to the map of local labels. It also enjoys the same benefits. The FCN-based face spoofing detection methods are generally superior to the CNN-based ones since they make full use of the basic properties of ubiquitous and repetitive. But the current FCN-based methods still have two limitations. First, since the input size

of the network is fixed, the face crops are resized from arbitrary size to fixed size in the preprocessing step. The image resizing will reduce the intensity of Moiré patterns in the high-frequency band and make the subsequent classification hard. Second, the FCN-based method cannot eliminate the domain shift which is an unsolved problem and promising research direction. Based on our previous FCN-based study [10], we are going to proposed a new method called Fully Convolutional Network with Domain Adaptation and Lossless Size Adaptation (FCN-DA-LSA).

E. DOMAIN ADAPTATION-BASED FACE SPOOFING DETECTION

Domain adaptation is a kind of transfer learning which handles multi-distribution data. It transfers knowledge from the source domain to the target domain. Most of them are based on deep feature manipulation. Some methods learn domain-invariant features by minimizing the distribution of feature across domains [37]–[41], maximizing the difference between the private features and the shared features [42], and learning adversarially against the domain classifier [43], [44]. The domain-invariant features makes whole classification invariant to domain changes. Some other methods [45], [46] concatenate the data/features in different domains, and encourage the subsequent classifiers to recognize the domains.

Since there are various cameras, lighting conditions, attack types, etc. across different datasets, face spoofing detection performance degrades when the training and testing datasets are different. Domain adaptation is an promising research direction in face spoofing detection. Yang *et al.* [29] learn person-specific face spoofing detectors for each subject domain. Li *et al.* [18] introduce the Maximum Mean Discrepancy (MMD) loss to face spoofing detection. Shao *et al.* [20] learn a generalized feature space by adversarial learning. The proposed domain adaptation layer is an extension of [46] which has never been used in face spoofing detection nor deep learning. Our augmented feature is a combination of shared features, source private features, and target private features. And the following neural network weights is a combination of shared weights, source private weights, and target private weights. The idea of modeling the private feature is similar to [42], [45].

F. TRADITIONAL FRUSTRATINGLY EASY DOMAIN ADAPTATION

The frustratingly easy domain adaptation [46] is closely related to this work. The original method is devised for traditional linear and kernel models. And we extend it to deep learning. Let's review the original linear and kernel methods first. To adapt the source domain D^s and the target domain D^t, the original algorithm augments the input $\boldsymbol{x} \in D^s \cup D^t \in \mathbb{R}^C$ by

$$F(\boldsymbol{x}) = \begin{cases} [\boldsymbol{x}, \boldsymbol{x}, \boldsymbol{0}] & \boldsymbol{x} \in D^s \\ [\boldsymbol{x}, \boldsymbol{0}, \boldsymbol{x}] & \boldsymbol{x} \in D^t \end{cases} \tag{1}$$

where $[\cdot, \cdot, \cdot]$ is the vector concatenation operator, $\boldsymbol{0} = (0, 0, \cdots, 0) \in \mathbb{R}^C$ and $F(\boldsymbol{x}) \in \mathbb{R}^{3C}$. The augmentation has a kernelized version. Let ϕ be the mapping from the input space to the Reproducing Kernel Hilbert Space (RKHS), and $k(\boldsymbol{x}, \boldsymbol{x}') = \langle \phi(\boldsymbol{x}), \phi(\boldsymbol{x}') \rangle$ be the kernel function. The augmentation in the RKHS (perhaps infinite-dimensional) is

$$F(\boldsymbol{x}) = \begin{cases} [\phi(\boldsymbol{x}), \phi(\boldsymbol{x}), \boldsymbol{0}] & \boldsymbol{x} \in D^s \\ [\phi(\boldsymbol{x}), \boldsymbol{0}, \phi(\boldsymbol{x})] & \boldsymbol{x} \in D^t \end{cases} \tag{2}$$

Since the RKHS is expanded, the new kernel function after augmentation is

$$K(\boldsymbol{x}, \boldsymbol{x}') = \begin{cases} 2k(\boldsymbol{x}, \boldsymbol{x}') & \textit{same domain} \\ k(\boldsymbol{x}, \boldsymbol{x}') & \textit{different domains} \end{cases} \tag{3}$$

III. METHODOLOGY

As its name suggests, the FCN-DA-LSA includes a lossless size adaptation preprocessor followed by an FCN-based pixel-level classifier embedded with a domain adaptation layer. The FCN-DA-LSA is divided into three individual parts of FCN, DA, and LSA which will be respectively elaborated in the following three subsections.

A. PIXEL-LEVEL CLASSIFICATION FCN BACKBONE

The depth estimating subnetwork proposed in [7] and the pixel-level classification network proposed in [10] share the same main architecture. To avoid create deep network arbitrarily and focus the attention on our contribution, we adopted this network as our backbone. Similar to the most widely used CNNs, the input of the FCN is a $256 \times 256 \times 3$ RGB image. The network contains thirteen 3×3 convolutional layers and three max-pooling layers. While the layer goes deeper, the spatial size of the feature maps gradually decreases to 32×32 pixels. Two short connections are placed in the middle of the network to encourage the network to learn features in different scales. The final convolutional layer is followed by sigmoid normalization rather than fully connected layers. The total number of the trainable parameters is 2.2M which is much smaller than most traditional CNNs with fully connected layers. The depth and convolutional kernel size of the network are close to the ones in 19-layered VGG [47]. The face spoofing detection does not need a very deep network since it is a low-level image task rather than a high-level semantic task. The two skip connections in the FCN provide extra flexibility for determining the network depth.

Pixel-wise cross-entropy loss is applied for optimization:

$$\hat{Y} = fcn_\theta(X), \tag{4}$$

$$l(X, y, \theta) = -\sum_{i=1}^{W}\sum_{j=1}^{H}\left(y \log(\hat{Y}_{i,j}) + (1-y)\log(1-\hat{Y}_{i,j})\right), \tag{5}$$

$$\hat{\theta} = \arg\min_\theta \sum_{X,y \in train\ set} l(X, y, \theta), \tag{6}$$

where X is the image, $y \in \{0, 1\}$ is the ground-truth label, and fcn_θ is the FCN with parameter θ. The loss l compares

the $W \times H$ FCN's prediction map with the ground-truth label y pixel by pixel. Finally, $\hat{\theta}$ is learned by minimizing the empirical total loss over the training set.

The main difference between the FCN and the common CNN is the prediction and the supervision labels. In face spoofing detection task, the decision can be sufficiently made when only a small local region is given, since the spoof clues, including Moiré patterns, certain LBP patterns, and compression/color degradation, etc. exist in every patch. The FCN is suitable for such tasks. But, this property does not generally exist in all the image classification tasks. For example, we cannot determine the identity from only a small local region of the face. To this end, the CNN is suitable for the face recognition task. We believe that the FCN is applicable for the face spoofing detection task for two reasons. First, the pixel-level local supervision is stronger than the image-level global supervision during training. It introduces similar effects of patch-based learning in an efficient way. Second, the decisions at different locations can be fused to further improve the accuracy during testing.

TABLE 1. The configuration details of the FCN.

Layer	# of activations	# of parameters
Input	$256 \times 256 \times 3$	0
Convolutional	$256 \times 256 \times 64$	$3 \times 3 \times 3 \times 64 + 64$
Convolutional	$256 \times 256 \times 128$	$3 \times 3 \times 64 \times 128 + 128$
Convolutional	$256 \times 256 \times 196$	$3 \times 3 \times 128 \times 196 + 196$
Convolutional	$256 \times 256 \times 128$	$3 \times 3 \times 196 \times 128 + 128$
Max-pooling	$128 \times 128 \times 128$	0
Convolutional	$128 \times 128 \times 128$	$3 \times 3 \times 128 \times 128 + 128$
Convolutional	$128 \times 128 \times 196$	$3 \times 3 \times 128 \times 196 + 196$
Convolutional	$128 \times 128 \times 128$	$3 \times 3 \times 196 \times 128 + 128$
Max-pooling	$64 \times 64 \times 128$	0
Convolutional	$64 \times 64 \times 128$	$3 \times 3 \times 128 \times 128 + 128$
Convolutional	$64 \times 64 \times 196$	$3 \times 3 \times 128 \times 196 + 196$
Convolutional	$64 \times 64 \times 128$	$3 \times 3 \times 196 \times 128 + 128$
Max-pooling	$32 \times 32 \times 128$	0
Downsample & concatenate	$32 \times 32 \times 384$	0
Domain adaptation	$32 \times 32 \times 768$	0
Convolutional	$32 \times 32 \times 128$	$3 \times 3 \times 384 \times 128 + 128$
Convolutional	$32 \times 32 \times 64$	$3 \times 3 \times 128 \times 64 + 64$
Convolutional	$32 \times 32 \times 1$	$3 \times 3 \times 64 \times 1 + 1$
Pixel-wise sigmoid normalization	$32 \times 32 \times 1$	0
Total	–	$2,245,133$

B. DEEP FEATURE AUGMENTATION-BASED DOMAIN ADAPTATION FOR DEEP NEURAL NETWORKS

In Fig. 3-(a), the domain adaptation layer embedded in the network is a deep-neural-networks-oriented extension of the frustratingly easy domain adaptation [46]. Based on the definitions of the original linear and kernel versions in Section II-F, it is straightforward to further derive a deep neural network version of the above approach. For fully connected neural networks, Let f be the non-linear mapping from the input $\boldsymbol{x}$ to a deep feature $f(\boldsymbol{x}) \in \mathbb{R}^C$, the augmentation in the deep feature space is

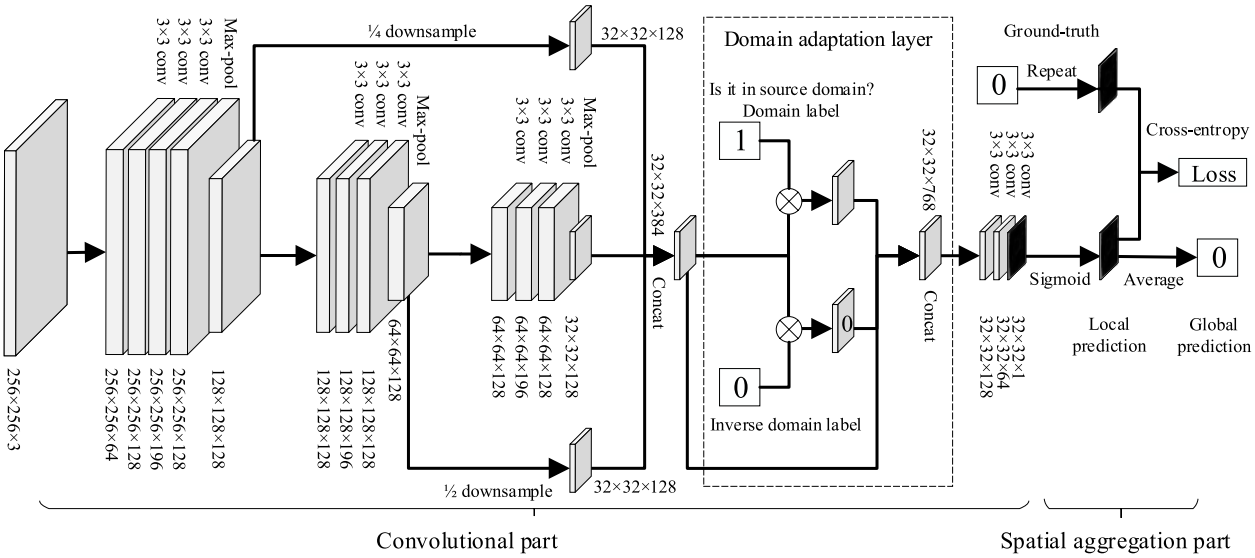
$$F(\boldsymbol{x}) = \begin{cases} [f(\boldsymbol{x}), f(\boldsymbol{x}), \mathbf{0}] & \boldsymbol{x} \in D^s \\ [f(\boldsymbol{x}), \mathbf{0}, f(\boldsymbol{x})] & \boldsymbol{x} \in D^t \end{cases} \tag{7}$$

For CNNs/FCNs, let $f(\boldsymbol{x}) \in \mathbb{R}^{W \times H \times C}$ be C deep feature maps of $W \times H$. $F(\boldsymbol{x}) \in \mathbb{R}^{W \times H \times 3C}$ is obtained by extending the $[\cdot, \cdot, \cdot]$ to the deep feature map concatenation which keep the spatial structure. The Eq. (1), Eq. (2), and Eq. (7) show the linear, kernelized, and the deep neural network version of the deep feature augmentation-based domain adaptation, respectively. They can be applied to a K-domain problem by simply expanding the deep feature space to $\mathbb{R}^{(K+1)C}$ where C is the space dimension before expansion.

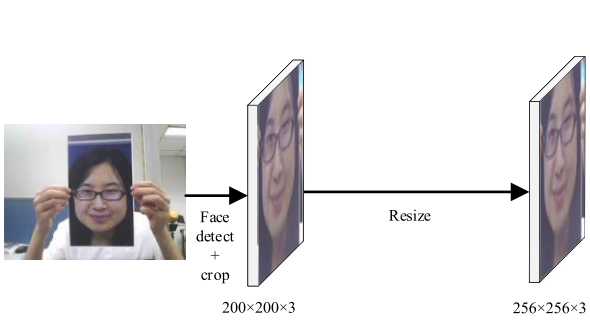
The domain adaptation layer can be defined based on Eq. (7). It can be inserted to any sequential deep neural networks to augment the deep features/feature maps by three times. Since the domain adaptation layer is derivable and parameterless, the whole network can be trained in an end-to-end scheme. It is worth noting that, unlike some domain adaptation methods [37]–[41] which learn domain-invariant features, our augmented feature is a combination of shared features, source private features, and target private features. And the following fully connected/convolutional weights is a combination of shared weights, source private weights, and target private weights. In Fig. 3-(a), the domain adaptation layer is inserted in the middle of the network since we need a stack of following layers to learn a non-linear classifier. The network extracts the convolutional features in multi-resolutions, expands and adapts the features, classifies the features pixel by pixel, finally fuses the decisions spatially.

C. LOSSLESS SIZE ADAPTATION

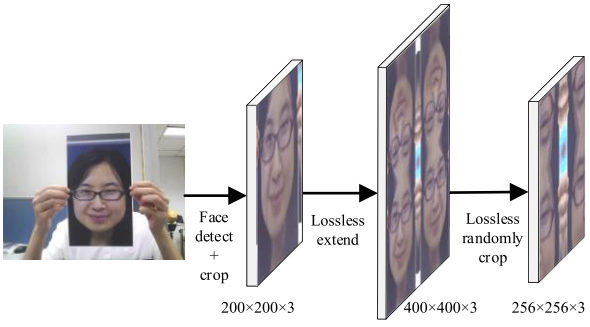
In photo and video attacks, spoofing clues (such as Moiré patterns, certain LBP patterns, and compression/color degradation) exist in the face area. As demonstrate in Fig. 3-(b), using face detector to crop faces from various backgrounds and resizing the cropped faces to fixed spatial size are common practices. Since image resizing will reduce the intensity of Moiré patterns in the high-frequency band and make the subsequent classification hard, Atoum *et al.* [8] use unresized face patches as the input of their patch-based CNN to avoid resizing to the original images and maintain the spoof patterns. But they only use ten random patches during training and testing which is not enough for a high-performance classification. Motivated by Atoum *et al.* [8], we use more implicit patches in an efficient fully convolutional scheme which fully utilizes the basic properties of ubiquitous and repetitive. Since the pixel-level activations are shared between adjacent sliding windows, thousands of implicit patches can be computed with acceptable time and space costs. To further preserve the original image scale, we design an lossless size adaptation preprocessor for the FCN. As demonstrate in Fig. 3-(c), after the face is detected and cropped, if the image size (e.g. 200×200 pixels) is smaller than the FCN's input size (e.g. 256×256 pixels), the image must be extended first. The extension on both x/y-axis will be repeated until the width/height is greater than or equal



(a) The FCN and the embedded Domain Adaptation (DA) layer. The network extracts the convolutional features in multi-resolutions, expands and adapts the features, classifies the features pixel by pixel, finally fuses the decisions spatially.



(b) The common image resizing preprocessor. It detects and crops faces, then resizes the faces to fit FCN's input.



(c) The Lossless Size Adaptation (LSA) preprocessor. It detects and crops faces, then extends faces to get sufficient pixels, finally randomly crops the extended images to fit FCN's input.

FIGURE 3. The pipeline of the proposed method.

to the FCN's input width/height. Since the face images are non-stationary signals in nature, the common zero-padding and periodic extension will bring discontinuous pixels in the image boundaries. To avoid additional sharp edges and keep the extended image smooth, we suggest using the symmetric extension which is wildly used in LBP-based face analysis [48], [49]. Then, a sub-image is randomly cropped to meet the exact size of the FCN's input. N color images in different sizes are spatially normalized to $W \times H$ pixels and concatenated into a tensor of $N \times H \times W \times 3$ which is compatible with modern deep-learning frameworks based on mini-batch gradient descent and parallel computing. The extension and the subsequent cropping is based on the assumption of the two basic properties of ubiquitous and repetitive which is discussed in Section II-D. Such a four-step (detection, cropping, extension, and cropping) size adaptation preprocessor can keep the FCN's input in its original scale, and preserve all the original high-frequency band information in it. It is essential for the performance of the whole system.

There are two alternative size adaptation schemes for testing the images of arbitrary spatial sizes. The first one is to use the same image sampling method during training: cropping one or more sub-images on the extended images, then fusing the decisions of all the crops to get a convincing result. For example, the Alex-Net [50] extracts four corner crops and one center crop. For high-dimensional faces (e.g. some 850 × 720 faces in the CASIA-FASD dataset [51]) the above scheme ignores useful information outside the crops. For low-dimensional faces (e.g. some 81 × 81 faces in the CASIA-FASD dataset) the extension makes it redundant and inefficient. The second size adaptation scheme is to directly feed the images of arbitrary spatial sizes into the convolutional layers, and adjust the spatial sizes of the activation maps correspondingly. Although the spatial size of the final activation map, i.e. the probabilities, changes every time, its average is still a good estimation of the whole face. The latter one will get better performance since it equally aggregates decisions at each location. Our experiment shows

TABLE 2. The amount of samples in the CASIA-FASD, Replay-Attack, and OULU-NPU datasets.

Dataset	# of subjects	# of videos	# of frames	Face detection result	# of genuine faces	# of spoof faces
CASIA-FASD	50	600	111,027	Detected	26,824	83,961
				Rejected	45	197
Replay-Attack	50	1,300	347,498	Detected	112,500	233,331
				Rejected	0	1,667
OULU-NPU	55	4,950	661,905	Detected	130,785	529,942
				Rejected	1	1,177

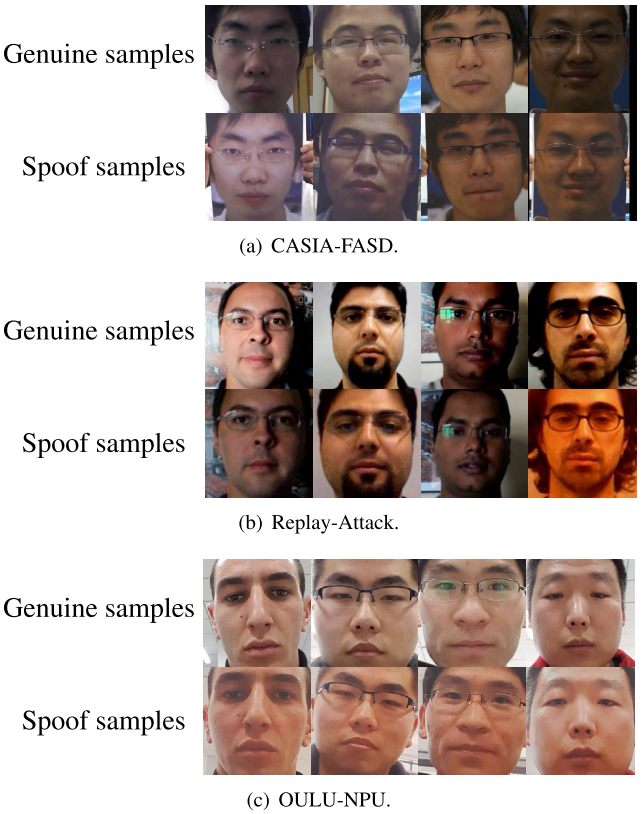
[1] A small portion ($\approx 0.28\%$) of the frames are rejected by the high-performance face detector. Most of them are spoof ones. This will have negligible effects on the video-level performance.

the advantage of the proposed lossless size adaptation preprocessor both in training and testing.

IV. EXPERIMENTS

A. DATASETS AND THE PREPROCESSING

Three datasets are used in the experiment, namely CASIA Face Anti-Spoofing Dataset (CASIA-FASD) [51], the Idiap Replay-Attack dataset [52], and the OULU-NPU dataset [53]. The CASIA-FASD dataset has 600 videos of 50 subjects. The Replay-Attack dataset has 1,300 videos of 50 subjects. And the OULU-NPU dataset has 4,950 videos of 55 subjects. The length of the video is about 6.5 seconds $\times$ 25 frame per second on average. A high-performance HOG+SVM based face detector [54] is used to extract a genuine face or a spoof face from each frame. A small portion ($\approx 0.28\%$) of the frames are rejected by the face detector. Most of them are spoof ones. This will have negligible effects on the video-level performance. Finally, we obtain 26,824/112,500/130,785 genuine faces and 83,961/233,331/529,942 spoof faces from the CASIA-FASD/Replay-Attack/OULU-NPU dataset, respectively. The sizes of the datasets before/after the face detection are listed in detail in Table 2. Some examples of the preprocessed faces are illustrated in Fig. 4.



(a) CASIA-FASD.

(b) Replay-Attack.

(c) OULU-NPU.

FIGURE 4. The examples of the preprocessed faces.

B. TRAINING AND TESTING PROTOCOLS

Let C and R be the abbreviation of the CASIA-FASD dataset and the Replay-Attack dataset, respectively. A short notation of "training set $\rightarrow$ testing set" is used to denote the training and testing protocols. For example, $C \rightarrow R$ and $R \rightarrow C$ are the commonly used cross-dataset protocols between CASIA-FASD and Replay-Attack. There are 50 subjects in each dataset. We choose 2/50 subjects from each dataset. The subsets are briefly notated by the initial letter C or R followed by a numeric subscript denoting the number of subjects in it. The subsets of the two subjects, namely C_2 and R_2, are used as the small-sample training data in the target domain. More exactly, the domain adaptation method uses the $C+R_2 \rightarrow R_{48}$ and $R + C_2 \rightarrow C_{48}$ protocols which are close to the $C \rightarrow R$ and $R \rightarrow C$ protocols. Protocols of $C \rightarrow R_{48}$, $R_2 \rightarrow R_{48}$, $C + R_2 \rightarrow R_{48}$, $R \rightarrow C_{48}$, $C_2 \rightarrow C_{48}$, and $R + C_2 \rightarrow C_{48}$ will be used in the ablation studies in Section IV-C. Protocols of $C \rightarrow R$ and $R \rightarrow C$ will be used for comparing with the existing methods in Section IV-D.

The OULU-NPU dataset provide four standard protocols. The protocol I, II, and III are cross-backgrounds, cross-Presentation-Attack-instrument (cross-PAI), and cross-camera protocols, respectively. By combining these protocols, we get the most challenging protocol IV which is very close to a cross-dataset protocol. The four standard protocols are followed for comparing with the existing methods in Section IV-D. 1/20 subjects are redivided from the testing set for training the domain adaptation models.

During training, the mini-batch Stochastic Gradient Descent (SGD) with a learning rate of 0.001 and a mini-batch size of 10 is employed to train the networks. According to Table 2, The ratio between the number of genuine and spoof faces is about 1:3, 1:2, and 1:4 in CASIA-FASD, Replay-Attack, and OULU-NPU, respectively. The class

TABLE 3. The HTER (%) of ablation study on CASIA-FASD dataset and Replay-Attack dataset.

Method	Cross $C \rightarrow R_{48}$	Intra $R_2 \rightarrow R_{48}$	Hybrid $C+R_2 \rightarrow R_{48}$	Cross $R \rightarrow C_{48}$	Intra $C_2 \rightarrow C_{48}$	Hybrid $R+C_2 \rightarrow C_{48}$
FCN	28.37	13.46	16.35	36.46	26.56	28.13
FCN-LSA [1]	26.52	13.14	15.95	36.63	25.35	27.95
FCN-DA [2]	–	–	11.70	–	–	22.61
FCN-DA-LSA [1,2]	–	–	**11.22**	–	–	**21.92**

[1] LSA: Lossless size adaptation.
[2] DA: Domain adaptation with the help of small-sample external training data in the target domain (2/50 subjects).

TABLE 4. The HTER (%) of cross-dataset evaluation on CASIA-FASD dataset and Replay-Attack dataset without domain-adaptation.

Method	$C \rightarrow R$	$R \rightarrow C$
Motion [57]	50.20	47.90
LBP [57]	55.90	57.60
LBP-TOP [57]	49.70	60.60
Motion-Mag [58]	50.10	47.00
Spectral Temporal Cubes [31]	34.40	50.00
CNN [59]	48.50	45.50
IQM+IQA+SVM [15]–[17]	37.35	40.90
Colour Texture 1 [13]	47.00	39.60
Colour Texture 2 [14]	30.30	37.70
Auxiliary [7]	27.60	**28.40**
Noise Model [11]	28.50	41.10
STASN [12]	31.50	30.90
FCN-LSA [1]	**27.31**	37.33

[1] LSA: Lossless size adaptation.

TABLE 5. The HTER (%) of cross-dataset evaluation on CASIA-FASD dataset and Replay-Attack dataset with domain-adaptation.

Type	Method	$C \rightarrow R$	$R \rightarrow C$
Few-shot	KSA+DA+SVM [18]	27.40	36.00
Few-shot	MMD-AAE [19]	31.58	44.59
Zero-shot	MADDoG [20]	22.19	24.50
Few-shot	**FCN-DA-LSA [1,2]**	**11.23**	**21.83**

[1] LSA: Lossless size adaptation.
[2] DA: Domain adaptation with the help of small-sample external training data in the target domain (2/50 subjects).

distributions of the datasets are imbalanced. To handle this problem, a light-weight random sampling method is employed [55]. More specifically, the training set is shuffled once before training and then divided into groups of genuine and spoof. In each iteration, a mini-batch is composed of five positive and five negative samples which are sequentially drawn from the genuine and spoof groups, respectively. The training is stopped after 400,000 iterations. During testing, we fix the trainable parameters and predict the frame-level probabilities first. By following the recent studies [7], [9], [11], [56], the frame-level probabilities predicted by neural networks are temporally averaged to get the better video-level decisions. Finally, the Half Total Error Rate (HTER) are evaluated and reported for the CASIA-FASD and Replay-Attack. The Average Classification Error Rate (ACER), Attack Presentation Classification Error Rate (APCER), and Bonafide Presentation Classification Error Rate (BPCER) are evaluated and reported for the OULU-NPU.

In Section IV-C, an ablation study about four related methods including FCN, FCN-LSA, FCN-DA, and FCN-DA-LSA will be conducted under the cross-dataset protocol. And then in Section IV-D, the proposed method will be further compared with 21 existing methods including DeepPixBiS [9], Auxiliary Supervision [7], Noise Model [11], STASN [12], Colour Texture [13], [14], IQM+IQA+SVM [15]–[17], KSA+DA+SVM [18], MMD-AAE [19], and MADDoG [20] under the cross-dataset protocols and the standard protocols of OULU-NPU.

C. ABLATION STUDY ABOUT FOUR RELATED NETWORKS

Since the Lossless Size Adaptation (LSA) works in the preprocessing stage and the Domain Adaptation (DA) works in the inferring stage, the DA and the LSA are two independent improvements based on the FCN. In this ablation study, four related networks, namely FCN, FCN-LSA, FCN-DA, and FCN-DA-LSA, are created and compared under the protocols of $C \rightarrow R_{48}$, $R_2 \rightarrow R_{48}$, $C + R_2 \rightarrow R_{48}$, $R \rightarrow C_{48}$, $C_2 \rightarrow C_{48}$, and $R + C_2 \rightarrow C_{48}$ to show the advantage of the DA and the LSA. The results in Table 3 are divided into two horizontal groups in which the testing sets are the same (left group: R_{48}, right group: C_{48}). The results can be compared.

Analysis of the DA layer: in the left group of Table 3, FCN with/without LSA obtains HTERs of 26.52%/28.37% under the cross-dataset protocol $C \rightarrow R_{48}$, and obtains HTERs of 13.14%/13.46% under the intra-dataset protocol $R_2 \rightarrow R_{48}$. A new hybrid protocol $C+R_2 \rightarrow R_{48}$ is defined by merging the training set of the above two protocols. Usually, the testing error will decrease once the training data are augmented. But if the domain of the augmented training data is different, the domain shift will poison the classifier and increase the testing error by $\approx 2.85\%$ (2nd column vs. 3rd column). The domain adaptation is designed to make full use of the cross-domain data. While the domain adaptation is introduced, the testing error will decrease by $\approx 1.84\%$ (2nd column vs. 3rd column). With the help of small-sample external training data in the target domain, the two domain adaptation-based network obtained the best results. The similar result can also be observed in the right group of Table 3 by switching the training and testing data. The domain adaptation maintains the dominant role in this ablation study.

TABLE 6. The performance under four standard protocols of OULU-NPU dataset.

Protocol	Method	ACER (%)	APCER (%)	BPCER (%)
I	LBP-SVM [9]	32.29	12.92	51.67
	IQM-SVM [17]	25.00	19.17	30.83
	CPqD [56]	6.90	2.90	10.80
	GRADIANT [56]	6.90	1.30	12.50
	Auxiliary [7]	1.60	1.60	1.60
	MILHP [60]	4.60	8.30	0.80
	STASN [12]	1.90	1.20	2.50
	Noise Model [11]	1.50	1.20	1.70
	DeepPixBiS [9]	**0.42**	0.83	0.00
	FCN+LSA [1]	**0.42**	0.00	0.83
	FCN+DA+LSA [1,2]	**0.42**	0.00	0.83
II	LBP-SVM [9]	25.14	30.00	20.28
	IQM-SVM [17]	14.72	12.50	16.94
	MixedFASNet [56]	6.10	9.70	2.50
	GRADIANT [56]	2.50	3.10	1.90
	Auxiliary [7]	2.70	2.70	2.70
	MILHP [60]	5.40	5.60	5.30
	STASN [12]	2.20	4.20	0.30
	Noise Model [11]	4.30	4.20	4.40
	DeepPixBiS [9]	5.97	11.39	0.56
	FCN+LSA [1]	2.78	3.33	2.22
	FCN+DA+LSA [1,2]	**1.95**	2.22	1.67
III	LBP-SVM [9]	25.92±11.25	28.50±23.05	23.33±17.98
	IQM-SVM [17]	21.95±8.09	21.94±9.99	21.95±16.79
	MixedFASNet [56]	6.50±4.60	5.30±6.70	7.80±5.50
	GRADIANT [56]	3.80±2.40	2.06±3.90	5.00±5.30
	Auxiliary [7]	2.90±1.50	2.70±1.30	3.10±1.70
	MILHP [60]	4.00±2.90	1.50±1.20	6.40±6.60
	STASN [12]	2.80±1.60	4.70±3.90	0.90±1.20
	Noise Model [11]	3.60±1.60	4.00±1.80	3.80±1.20
	DeepPixBiS [9]	11.11±9.40	11.67±19.57	10.56±14.06
	FCN+LSA [1]	2.92±1.34	3.89±2.93	1.94±1.74
	FCN+DA+LSA [1,2]	**2.09±0.98**	2.78±1.76	1.39±1.12
IV	LBP-SVM [9]	48.33±6.07	41.67±27.03	55.00±21.21
	IQM-SVM [17]	36.67±12.13	34.17±25.89	39.17±23.35
	Massy HNU [56]	22.10±17.60	35.80±35.30	8.30±4.10
	GRADIANT [56]	10.00±5.00	5.00±4.50	15.00±7.10
	Auxiliary [7]	9.50±6.00	9.30±5.60	10.40±6.00
	MILHP [60]	12.00±6.20	15.80±12.80	8.30±15.70
	STASN [12]	7.50±4.70	6.70±10.60	8.30±8.40
	Noise Model [11]	5.60±5.70	5.10±6.30	6.10±5.10
	DeepPixBiS [9]	25.00±12.67	36.67±29.67	13.33±16.75
	FCN+LSA [1]	9.45±5.04	11.39±6.90	7.50±6.30
	FCN+DA+LSA [1,2]	**5.56±2.52**	6.94±3.95	4.17±3.15

[1] LSA: Lossless size adaptation.
[2] DA: Domain adaptation with the help of small-sample external training data in the target domain (1/20 subjects).

Analysis of the LSA preprocessor: as listed in Table 3, the FCN-LSA/FCN-DA-LSA generally outperform the FCN/FCN-DA for each protocol. The LSA can further decrease the HTER by $\approx 0.62\%$ (1st row vs. 2nd row, 3rd row vs. 4th row). It slightly improves the performance by keeping the original patterns in the preprocessing step, and plays a secondary role in this ablation study.

By combining the above two factors of DA and LSA, the FCN-DA-LSA is the best method for face spoofing detection in this ablation study. As listed in Table 3, with the help of small-sample external training data in the target domain, the FCN-DA-LSA obtains HTER of 11.22% and 21.92% under two hybrid protocols, respectively. The margins between the performances of the basic FCN and the improved FCN-DA-LSA is about 5.67% (1st row vs. 4th row). This is mainly achieved by making full use of the small-sample external training data.

D. COMPARISON WITH THE EXISTING METHODS

The proposed FCN-LSA and FCN-DA-LSA are further compared with 21 existing methods including DeepPixBiS [9], Auxiliary Supervision [7], Noise Model [11], STASN [12], Colour Texture [13], [14], IQM+IQA+SVM [15]–[17], KSA+DA+SVM [18], MMD-AAE [19], and MADDoG [20] under the cross-dataset protocols on CASIA-FASD and Replay-Attack and the four standard OULU-NPU protocols. Two conclusions can be made according to the results listed in Table 4, 5 and 6. First, the basic FCN-LSA achieves competitive performance among the compared ones. It wins the 1/13 and 3/13 places under the two cross-dataset protocols in Table 4. It also wins the 1/10, 4/10, 3/10, and 3/10 places under the four OULU-NPU protocols in Table 6. Second, with the help of small-sample external training data in the target domain, the FCN-DA-LSA outperforms three domain adaptation based face spoofing detection methods. It is currently the best under all six protocols in Table 5 and 6. The domain adaptation method extremely improves the performance. If the domain shift is large enough, e.g. under the two cross-dataset protocols and OULU-NPU protocol IV, the domain adaptation almost halve the errors (from 27.31% to 11.23% in protocol $C \rightarrow R$, from 37.33% to 21.83% in protocol $R \rightarrow C$, and from 9.45% to 5.56% in OULU-NPU protocol IV). The FCN-LSA and FCN-DA-LSA generalize well across different domains.

V. CONCLUSION

The FCN-DA-LSA is proposed for face spoofing detection in this paper. Its improvements are twofold. First, the Domain Adaptation method makes full use of the small-sample external training data in the target domain. Second, the Lossless Size Adaptation method preserves the high-frequent spoof clues caused by the face recapture process. The proposed method is empirically verified in the ablation study and also compared with the existing methods. The FCN-LSA obtains competitive performance among the state-of-the-art methods. With the help of small-sample external training data in the target domain, the FCN-DA-LSA further improves the performance and outperforms the existing methods.

The proposed deep feature augmentation is a kind of supervised few-shot domain adaptation. It can be employed when the target domain is already known. The requirement of external data is its major limitation. For example, in the cross-PAI or the cross-camera experiments, performance can be greatly improved (almost halve the error in the experiment) once fewer data (only one subjects in the experiment) in the target domain are given. In the future, we can explore some unsupervised few-shot domain adaptation methods or some zero-shot learning methods to relieve the external data limitation.

REFERENCES

[1] F. Deeba, H. Memon, F. Ali, A. Ahmed, and A. Ghaffar, "LBPH-basedenhanced real-time face recognition," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 5, pp. 274–280, 2019.

[2] K. Arya, S. S. Rajput, and S. Upadhyay, "Noise-robust low-resolution face recognition using sift features," in *Computational Intelligence: Theories, Applications and Future Directions*, vol. 2. New York, NY, USA: Springer, 2019, pp. 645–655.

[3] A. Ahmed, J. Guo, F. Ali, F. Deeba, and A. Ahmed, "LBPH based improved face recognition at low resolution," in *Proc. Int. Conf. Artif. Intell. Big Data (ICAIBD)*, May 2018, pp. 144–147.

[4] D. Menotti, G. Chiachia, A. Pinto, W. Robson Schwartz, H. Pedrini, A. Xavier Falcao, and A. Rocha, "Deep representations for iris, face, and fingerprint spoofing detection," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 4, pp. 864–879, Apr. 2015.

[5] Y. A. Ur Rehman, L. M. Po, and M. Liu, "Deep learning for face anti-spoofing: An end-to-end approach," in *Proc. Signal Process., Algorithms, Archit., Arrangements, Appl. (SPA)*, Sep. 2017, pp. 195–200.

[6] C. Nagpal and S. R. Dubey, "A performance evaluation of convolutional neural networks for face anti spoofing," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2019, pp. 1–8.

[7] Y. Liu, A. Jourabloo, and X. Liu, "Learning deep models for face anti-spoofing: Binary or auxiliary supervision," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 389–398.

[8] Y. Atoum, Y. Liu, A. Jourabloo, and X. Liu, "Face anti-spoofing using patch and depth-based CNNs," in *Proc. IEEE Int. Joint Conf. Biometrics (IJCB)*, Oct. 2017, pp. 319–328.

[9] A. George and S. Marcel, "Deep pixel-wise binary supervision for face presentation attack detection," in *Proc. Int. Conf. Biometrics (ICB)*, Jun. 2019, pp. 1–8.

[10] W. Sun, Y. Song, C. Chen, J. Huang, and A. C. Kot, "Face spoofing detection based on local ternary label supervision in fully convolutional networks," *IEEE Trans. Inf. Forensics Security*, early access, Apr. 3, 2020, doi: 10.1109/TIFS.2020.2985530.

[11] A. Jourabloo, Y. Liu, and X. Liu, "Face de-spoofing: Anti-spoofing via noise modeling," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 290–306.

[12] X. Yang, W. Luo, L. Bao, Y. Gao, D. Gong, S. Zheng, Z. Li, and W. Liu, "Face anti-spoofing: Model matters, so does data," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3507–3516.

[13] Z. Boulkenafet, J. Komulainen, and A. Hadid, "Face anti-spoofing based on color texture analysis," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2015, pp. 2636–2640.

[14] Z. Boulkenafet, J. Komulainen, and A. Hadid, "Face spoofing detection using colour texture analysis," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 8, pp. 1818–1830, Aug. 2016.

[15] O. Nikisins, A. Mohammadi, A. Anjos, and S. Marcel, "On effectiveness of anomaly detection approaches against unseen presentation attacks in face anti-spoofing," in *Proc. Int. Conf. Biometrics (ICB)*, Feb. 2018, pp. 75–81.

[16] D. Wen, H. Han, and A. K. Jain, "Face spoof detection with image distortion analysis," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 4, pp. 746–761, Apr. 2015.

[17] J. Galbally, S. Marcel, and J. Fierrez, "Image quality assessment for fake biometric detection: Application to iris, fingerprint, and face recognition," *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 710–724, Feb. 2014.

[18] H. Li, W. Li, H. Cao, S. Wang, F. Huang, and A. C. Kot, "Unsupervised domain adaptation for face anti-spoofing," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 7, pp. 1794–1809, Jul. 2018.

[19] H. Li, S. J. Pan, S. Wang, and A. C. Kot, "Domain generalization with adversarial feature learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 5400–5409.

[20] R. Shao, X. Lan, J. Li, and P. C. Yuen, "Multi-adversarial discriminative deep domain generalization for face presentation attack detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 10023–10031.

[21] K. Patel, H. Han, and A. K. Jain, "Secure face unlock: Spoof detection on smartphones," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 10, pp. 2268–2283, Oct. 2016.

[22] S. Kumar, S. Singh, and J. Kumar, "A comparative study on face spoofing attacks," in *Proc. Int. Conf. Comput., Commun. Autom. (ICCCA)*, May 2017, pp. 1104–1108.

[23] H. Muammar and P. L. Dragotti, "An investigation into aliasing in images recaptured from an LCD monitor using a digital camera," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2013, pp. 2242–2246.

[24] D. Caetano Garcia and R. L. de Queiroz, "Face-spoofing 2D-detection based on Moiré-pattern analysis," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 4, pp. 778–786, Apr. 2015.

[25] R. Ni, Y. Zhao, and X. Zhai, "Recaptured images forensics based on color moments and DCT coefficients features," *J. Inf. Hiding Multimedia Signal Process.*, vol. 6, no. 2, pp. 323–333, 2015.

[26] J. Maatta, A. Hadid, and M. Pietikainen, "Face spoofing detection from single images using micro-texture analysis," in *Proc. Int. Joint Conf. Biometrics (IJCB)*, Oct. 2011, pp. 1–7.

[27] K. Patel, H. Han, A. K. Jain, and G. Ott, "Live face video vs. Spoof face video: Use of moiré patterns to detect replay video attacks," in *Proc. Int. Conf. Biometrics (ICB)*, May 2015, pp. 98–105.

[28] S. R. Arashloo, J. Kittler, and W. Christmas, "Face spoofing detection based on multiple descriptor fusion using multiscale dynamic binarized statistical image features," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 11, pp. 2396–2407, Nov. 2015.

[29] J. Yang, Z. Lei, D. Yi, and S. Z. Li, "Person-specific face antispoofing with subject domain adaptation," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 4, pp. 797–809, Apr. 2015.

[30] T. D. Freitas Pereira, J. Komulainen, A. Anjos, J. M. De Martino, A. Hadid, M. Pietikäinen, and S. Marcel, "Face liveness detection using dynamic texture," *EURASIP J. Image Video Process.*, vol. 2014, no. 1, Dec. 2014.

[31] K. Simonyan and A. Zisserman, "Two-stream convolutional networks for action recognition in videos," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 568–576.

[32] W. Sun, H. Zhao, and Z. Jin, "3D convolutional neural networks for facial expression classification," in *Proc. Asian Conf. Comput. Vis.* New York, NY, USA: Springer, 2016, pp. 528–543.

[33] W. Sun, H. Zhao, and Z. Jin, "A facial expression recognition method based on ensemble of 3D convolutional neural networks," *Neural Comput. Appl.*, vol. 31, no. 7, pp. 2795–2812, Jul. 2019.

[34] W. Yin, Y. Ming, and L. Tian, "A face anti-spoofing method based on optical flow field," in *Proc. IEEE 13th Int. Conf. Signal Process. (ICSP)*, Nov. 2016, pp. 1333–1337.

[35] A. Pinto, H. Pedrini, W. R. Schwartz, and A. Rocha, "Face spoofing detection through visual codebooks of spectral temporal cubes," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 4726–4740, Dec. 2015.

[36] M. De Marsico, M. Nappi, D. Riccio, and J.-L. Dugelay, "Moving face spoofing detection via 3D projective invariants," in *Proc. 5th IAPR Int. Conf. Biometrics (ICB)*, Mar. 2012, pp. 73–78.

[37] A. Rozantsev, M. Salzmann, and P. Fua, "Beyond sharing weights for deep domain adaptation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 4, pp. 801–814, Apr. 2019.

[38] M. Long, H. Zhu, J. Wang, and M. I. Jordan, "Deep transfer learning with joint adaptation networks," in *Proc. 34th Int. Conf. Mach. Learn. JMLR*, vol. 70, Aug. 2017, pp. 2208–2217.

[39] B. Sun and K. Saenko, "Deep coral: Correlation alignment for deep domain adaptation," in *Proc. Eur. Conf. Comput. Vis.* New York, NY, USA: Springer, 2016, pp. 443–450.

[40] B. Sun, J. Feng, and K. Saenko, "Return of frustratingly easy domain adaptation," in *Proc. 13th AAAI Conf. Artif. Intell.*, 2016, pp. 2058–2065.

[41] X. Zhang, F. Xinnan Yu, S.-F. Chang, and S. Wang, "Deep transfer network: Unsupervised domain adaptation," 2015, *arXiv:1503.00591*. [Online]. Available: http://arxiv.org/abs/1503.00591

[42] K. Bousmalis, G. Trigeorgis, N. Silberman, D. Krishnan, and D. Erhan, "Domain separation networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 343–351.

[43] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, "Domain-adversarial training of neural networks," in *Domain Adaptation in Computer Vision Applications*. New York, NY, USA: Springer, 2017, pp. 189–209.

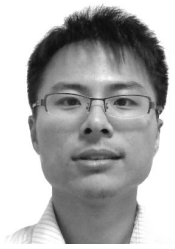
[44] E. Tzeng, J. Hoffman, N. Zhang, K. Saenko, and T. Darrell, "Deep domain confusion: Maximizing for domain invariance," 2014, *arXiv:1412.3474*. [Online]. Available: http://arxiv.org/abs/1412.3474

[45] S. Chopra, S. Balakrishnan, and R. Gopalan, "Dlid: Deep learning for domain adaptation by interpolating between domains," in *Proc. ICML Workshop Challenges Represent. Learn.*, vol. 2, no. 6, Jun. 2013.

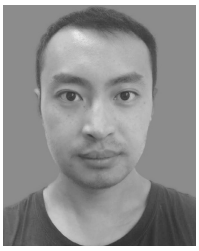
[46] H. Daumé III, "Frustratingly easy domain adaptation," in *Proc. 45th Annu. Meeting Assoc. Comput. Linguistics*, 2007, pp. 256–263.

[47] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Representations*, 2015.

[48] W.-L. Chao, J.-J. Ding, and J.-Z. Liu, "Facial expression recognition based on improved local binary pattern and class-regularized locality preserving projection," *Signal Process.*, vol. 117, pp. 1–10, Dec. 2015.
[49] K.-Y. Tsai, J.-J. Ding, and Y.-C. Lee, "Frontalization with adaptive exponentially-weighted average ensemble rule for deep learning based facial expression recognition," in *Proc. IEEE Asia–Pacific Conf. Circuits Syst. (APCCAS)*, Oct. 2018, pp. 447–450.
[50] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
[51] Z. Zhang, J. Yan, S. Liu, Z. Lei, D. Yi, and S. Z. Li, "A face antispoofing database with diverse attacks," in *Proc. 5th IAPR Int. Conf. Biometrics (ICB)*, Mar. 2012, pp. 26–31.
[52] I. Chingovska, A. Anjos, and S. Marcel, "On the effectiveness of local binary patterns in face anti-spoofing," in *Proc. Int. Conf. Biometrics Special Interest Group (BIOSIG)*, Sep. 2012, pp. 1–7.
[53] Z. Boulkenafet, J. Komulainen, L. Li, X. Feng, and A. Hadid, "OULU-NPU: A mobile face presentation attack database with real-world variations," in *Proc. 12th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, May 2017, pp. 612–618.
[54] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 1, Jun. 2005, pp. 886–893.
[55] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1263–1284, Sep. 2009.
[56] Z. Boulkenafet, J. Komulainen, Z. Akhtar, A. Benlamoudi, D. Samai, S. E. Bekhouche, A. Ouafi, F. Dornaika, A. Taleb-Ahmed, and L. Qin, "A competition on generalized software-based face presentation attack detection in mobile scenarios," in *Proc. IEEE Int. Joint Conf. Biometrics (IJCB)*, Oct. 2017, pp. 688–696.
[57] T. de Freitas Pereira, A. Anjos, J. M. De Martino, and S. Marcel, "Can face anti-spoofing countermeasures work in a real world scenario?" in *Proc. Int. Conf. Biometrics (ICB)*, Jun. 2013, pp. 1–8.
[58] S. Bharadwaj, T. I. Dhamecha, M. Vatsa, and R. Singh, "Computationally efficient face spoofing detection with motion magnification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2013, pp. 105–110.
[59] J. Yang, Z. Lei, and S. Z. Li, "Learn convolutional neural network for face anti-spoofing," 2014, *arXiv:1408.5601*. [Online]. Available: http://arxiv.org/abs/1408.5601
[60] C. Lin, Z. Liao, P. Zhou, J. Hu, and B. Ni, "Live face verification with multiple instantialized local homographic parameterization," in *Proc. 27Int. Joint Conf. Artif. Intell.*, Jul. 2018, pp. 814–820.



WENYUN SUN (Member, IEEE) received the B.S. degree in computer science and technology and the M.S. degree in pattern recognition and intelligent system from the Jiangsu University of Science and Technology, China, in 2009 and 2012, respectively, and the Ph.D. degree in pattern recognition and intelligent systems from the Nanjing University of Science and Technology, China, in 2018. He is currently a Postdoctoral Researcher with the Guangdong Key Laboratory of Intelligent Information Processing and Shenzhen Key Laboratory of Media Security, College of Electronics and Information Engineering, Shenzhen University, China. His current research interests are in the areas of deep learning and face analysis.



YU SONG (Member, IEEE) received the B.S. degree in information engineering and the M.S. and Ph.D. degrees in signal and information processing from the Nanjing University of Aeronautics and Astronautics, China, in 2010, 2013, and 2018, respectively. He is currently a Postdoctoral Researcher with the Guangdong Key Laboratory of Intelligent Information Processing and Shenzhen Key Laboratory of Media Security, College of Electronics and Information Engineering, Shenzhen University, China. His current research interests include deep learning and pattern recognition.

HAITAO ZHAO received the Ph.D. degree in pattern recognition and intelligent system from the Nanjing University of Science and Technology, Nanjing, China, in 2003. He is currently a Professor with the East China University of Science and Technology, Shanghai, China. His current interests are in the areas of pattern recognition, machine learning, and computer vision.

ZHONG JIN received the B.S. degree in mathematics, the M.S. degree in applied mathematics, and the Ph.D. degree in pattern recognition and intelligent system from the Nanjing University of Science and Technology, Nanjing, China, in 1982, 1984, and 1999, respectively. His current research interests are in the areas of pattern recognition and face recognition.